ADAPTIVELY WEIGHTED, PARTITIONED CONTEXT EDIT DISTANCE STRING MATCHING

ABSTRACT:

A system and method for examining a string of symbols and
identifying portions of the string which match a predetermined
pattern using adaptively weighted, partitioned context edit
distances. A pattern is partitioned into context and value
components, and candidate matches for each of the components
is identified by calculating an edit distance between that
component and each potentially matching set (sub-string) of
symbols within the string. One or more candidate matches
having the lowest edit distances are selected as matches for
the pattern. The weighting of each of the component matches
may be adapted to optimize the pattern matching and, in one
embodiment, the context components may be heavily weighted to
obtain matches of a value for which the corresponding pattern
is not well defined. In one embodiment, an edit distance
matrix is evaluated for each of a prefix component, a value
component and a suffix component of a pattern.   The
evaluation of the prefix matrix provides a basis for
identifying indicators of the beginning of a value window,
while the evaluation of the suffix matrix provides a basis for
identifying the alignment of the end of the value window. The
value within the value window can then be evaluated via the
value matrix to determine a corresponding value match score.

013018.00019:117937.01